

DAS PROJEKT PLIDIS

Problemlösendes Informationssystem mit Deutsch als
Interaktionssprache

Die Konzeption, nach der ein Informationssystem mit Deutsch als Interaktionssprache zu entwickeln ist, kann charakterisiert werden durch die zugrundegelegten Annahmen:

- (1) Es ist möglich, eine formale Sprache zu konstruieren (Konstruktsprache), die zwei Funktionen erfüllt:
 - Semantiksprache zur formalen Darstellung des Inhalts von natürlichsprachlichen Äußerungen wie Wörtern, Sätzen und Texten zu sein, sowie
 - Modellbeschreibungssprache zur Darstellung der Objekte, Beziehungen und Gesetzmäßigkeiten, die in einem "Weltausschnitt" (im Sinne von Sachbereich oder Problembereich) vorzufinden sind;
- (2) es ist möglich, Operationen über dieser Konstruktsprache zu definieren, die die Fähigkeit nachbilden, Sprache zu verstehen und Probleme zu lösen, wie etwa Verstehen als Auflösen von Mehrdeutigkeiten, Rückfragen, Folgerungen ziehen, Antworten;
- (3) es ist möglich, Regeln einer Übersetzungsgrammatik anzugeben, die beschreiben, wie natürlichsprachliche Äußerungen in diese Konstruktsprache zu übersetzen sind und wie konstrukt-sprachliche Formeln in natürliche Sprache zurückzuübersetzen sind;
- (4) es ist möglich, im Rahmen eingeschränkter Lösungen für (1) bis (3) sinnvolle Realisierungen auf einer DV-Anlage zu implementieren, d.h. Systeme, die nicht nur labormäßig Ausschnitte von Sprachverstehen und Problemlösung simulieren, sondern die auch in der Praxis erprobbar sind.

PLIDIS besteht aus

- einem linguistisch-logischen Teil, der die deutschsprachige Eingabe in eine Konstruktsprache überführt, in der die interne Repräsentation (IR) der Eingabe erfolgt;
- einem Problemlösungsteil, der die üblichen Speicher- und Retrievalaufgaben übernimmt und standardmäßige Problemlösungsoperationen zur Verfügung stellt.

Aus mehreren Gründen ist es erstrebenswert, das System so weit wie möglich modular zu gestalten. Zum einen wird hierdurch die Interaktion auf möglichst vielen Stufen einer Problembearbeitung möglich, was eine genauere Lokalisierung etwa auftretender Schwierigkeiten erleichtert. Zum anderen wird die Voraussetzung für eine effiziente Zusammenarbeit und Austestung des Systems und seiner Komponenten in der Gruppenarbeit geboten.

Die folgende Abbildung zeigt schematisch die Hauptkomponenten eines solchen Systems mit dem Datenfluß zwischen den einzelnen Komponenten.

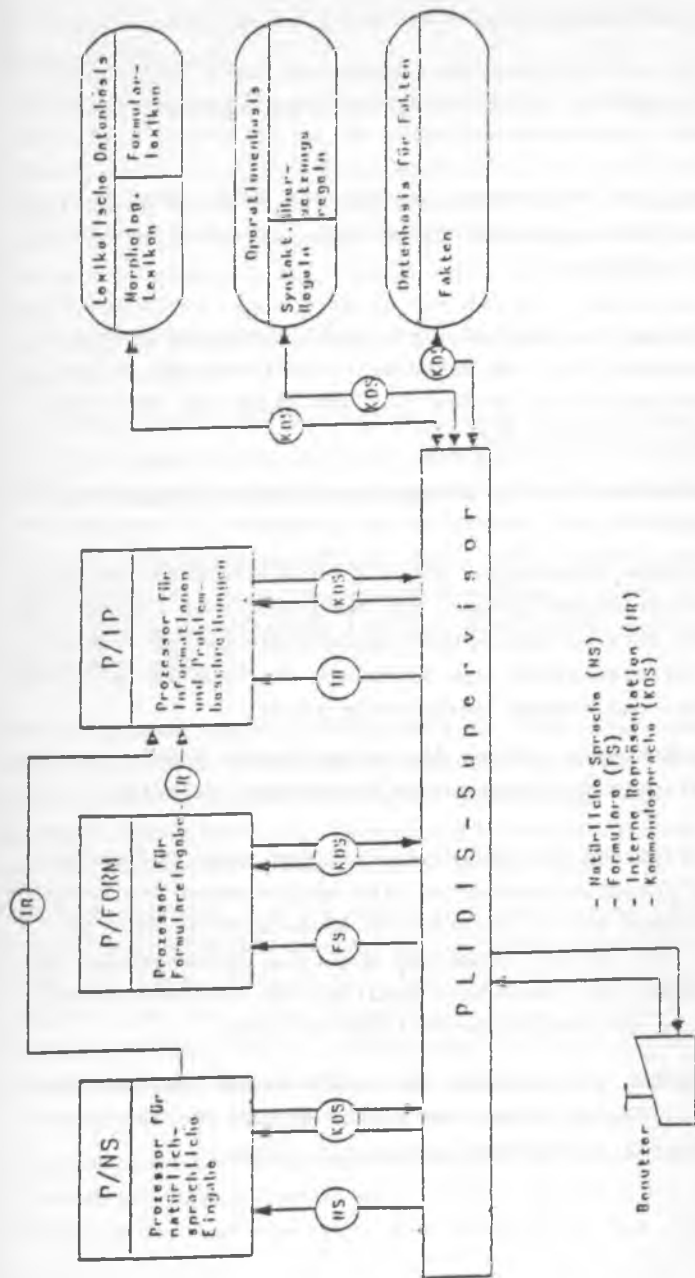


Abbildung: PLIDIS-Hauptkomponenten und Datenfluß

Erläuterungen zur Abbildung:

Der *Prozessor für natürlichsprachliche Eingaben* analysiert eingegebene Sätze und übersetzt sie in die PLIDIS-interne Repräsentationsform (IR).

Zur Verarbeitung von Massendaten stereotyper Form, deren Eingabe in natürlicher Sprache zu umständlich wäre, ist ein *Prozessor für Formulareingabe* verfügbar.

Die Interaktion der Komponenten wird vom *PLIDIS-Supervisor* gesteuert. Diese Steuerung kann vom Benutzer durch Kommandos in der Kommandosprache beeinflusst werden, was vor allem bei Test- und Wartungsaufgaben notwendig ist.

Die Systemkomponenten greifen entsprechend ihren Aufgaben auf die *Datenbasen* zu:

- eine lexikalische Datenbasis, die ein morphologisches Lexikon und ein Formularlexikon umfaßt; das morphologische Lexikon enthält zu jedem verarbeitbaren Wort morphologische Informationen, während im Formularlexikon die Formulare und Tabellen zur formatierten Ein- und Ausgabe verzeichnet sind;
- eine Operationenbasis, welche die syntaktischen Regeln und die Überführungsregeln für eingegebene Äußerungen enthalten.

Die komplexen Probleme der Sprachverarbeitung sowie der derzeitige Stand der Forschung erlaubten eine Realisierung des skizzierten PLIDIS-Konzepts nur in sukzessiven Ausbaustufen. Dies bedeutet einmal, daß aus der Menge der möglichen Anforderungen an ein ideales PLIDIS ein begrenzter qualitativer Leistungsumfang für konkrete Ausbaustufen ausgewählt werden mußte.

Andererseits mußten im Interesse der Realisierung des Gesamtkonzepts die Einzellösungen unter dem Gesichtspunkt der Erweiterbarkeit und Generalisierbarkeit entwickelt werden.

Im einzelnen ergeben sich aus dem PLIDIS-Konzept folgende Mindestanforderungen:

- Verarbeitung und Verständnis natürlicher Sprache: PLIDIS soll fachliche Kommunikation im gewählten Sachbereich ermöglichen. Grundsätzlich sollten Äußerungsformen, die Faktizität (also gesicherte Daten über einem eindeutig darstellbaren Sachbereich), Regularitäten und Normen betreffen, zugelassen sein. Mindestanforderung war deshalb, daß morphosyntaktisch und semantisch nicht nur Kurzsätze mit Subjekt, Prädikat, Objekt, sondern auch Ein-Satz-Texte fast beliebiger Komplexität verarbeitbar sind.
- Interne Darstellung von Informationen: Die interne Darstellung des Weltausschnitts muß eine vergleichbare konzeptuelle Komplexität wie das Sprechen in natürlicher Sprache über diesen Weltausschnitt zulassen; nur so können alle natürlichsprachlichen Informationen adäquat (d.h. unter Wahrung des Informationsgehalts) intern dargestellt werden. Darüber hinaus muß die interne Darstellung auch der Komplexität des Weltbereichs angepaßt sein.

Darüber hinaus müssen verschiedene arithmetische, mengentheoretische und statistische Operationen, Kontrollen, ob eingegebene Daten Normen erfüllen oder nicht, bewältigbar sein. Diese Anforderungen begründeten die Verwendung eines deduktiven Systems, das zu arithmetischen Operationen in der Lage ist.

- Geschwindigkeit: Die Interaktionen des Systems müssen ein der jeweiligen Situation angepaßtes Antwortzeitverhalten aufweisen. In vielen Fällen, z.B. beim Speichern von Massendaten, spielt hier die reale Verweilzeit keine besondere Rolle; in anderen Fällen jedoch, insbesondere bei einer Frage-Antwort-Interaktion, ist ein der Reaktionszeit eines menschlichen Sprechers angenähertes Antwortzeitverhalten anzustreben.
- Übertragbarkeit und Erweiterbarkeit: Die entwickelten Verfahren und Darstellungsformen sollen von der Form her so allgemein sein, daß sie nicht qualitativ verändert werden müssen,

wenn der Anwendungsbereich um qualitativ gleichartige Bereiche erweitert wird oder durch einen qualitativ gleichwertigen Bereich ersetzt wird. In einem solchen Fall müßten demnach, wenn die Forderung nach Übertragbarkeit erfüllt ist, lediglich die Datenbasis und das semantische Lexikon angepaßt werden. Alle anderen Teile, insbesondere die Struktur des Systems und die Formate der Datenbasen, würden sich gleichbleiben. Eine erweiterte Form der Erweiterbarkeitsforderung besteht darin, daß auch die nächsten absehbaren qualitativen Erweiterungsschritte ohne Änderung der Konzeption möglich sind.

Das oben beschriebene Informationssystem wurde auf ein kommerzielles Datenbanksystem - ADABAS - aufgesetzt, das die Verwaltung der realen Daten des Anwenders übernahm. Hierzu war es notwendig, eine Schnittstelle von PLIDIS zu ADABAS zu schaffen.

Durch die Übertragung der Datenverwaltung auf ein schon eingesetztes System - eine Eigenentwicklung war aus Personal- und Zeitgründen nicht vertretbar - kamen einige Einschränkungen in die ursprüngliche Konzeption des PLIDIS herein, die vorab nicht absehbar waren, in obiger Beschreibung jedoch berücksichtigt sind.

Die Beantwortung von Anfragen an die Datenbasis, die Informationen zur Abwasserüberwachung von Betrieben in Baden-Württemberg enthält, erfolgt mit dem PLIDIS-System, das sich dazu der folgenden Komponenten bedient:

(a) Morphologische Analyse

Die morphologische Analyse besteht lediglich darin, daß jeder Wortform des Eingabesatzes eine oder mehrere morphosyntaktische Beschreibungen zugeordnet werden. Diese Beschreibungen stehen in einem Vollformlexikon. Zusätzlich zu diesem morphologischen Vollformlexikon (MOLEX) wurde ein MOLEX-Wortformengenerator geschaffen, der es erlaubt, aus dem Lexem und einer Wortklassennummer sämtliche Vollformen MOLEX-gerecht zu

⁺ Vgl. dazu den folgenden Beitrag von M. Kolvenbach

generieren. [†]Mithilfe dieses MOLEX-Generators wurden die Daten des Anwenders für die morphologische Aufsuche aufbereitet.

(b) Syntaktische Analyse

Die der syntaktischen Analyse zugrundeliegende Grammatik des Deutschen ist in Anlehnung an die von Woods beschriebenen 'augmented transition networks' formuliert. Mit dieser Grammatik können fast alle Konstruktionen des Deutschen auf Satzebene erkannt werden.

(c) Überführung in die interne Repräsentation

Zur internen Repräsentation der Fragen und Aussagen wurde eine Konstruktsprache KS entwickelt, die sich an der Prädikatenlogik erster Stufe orientiert.

Die Repräsentation in der Konstruktsprache KS wurde in mehrere Zwischenebenen aufgespalten - nämlich Zwischen-KS, Ziel-KS und Datenbasis-KS, wobei für alle Sprachen der Zwischenebenen dieselbe Syntaxdefinition gilt. So überführen die Überführungsregeln von der natürlichen Sprache in die Zwischen-KS. Die Zwischen-KS ist der natürlichen Sprache enger angelehnt als die Ziel-KS. Der Schritt von Zwischen-KS zu Ziel-KS erfolgt über Bedeutungspostulate, wobei Ziel-KS die Konstruktsprache ist, die die Tiefenstruktur des gewählten Weltausschnitts widerspiegelt, im Gegensatz zu der eher oberflächenbezogenen Darstellung in Zwischen-KS. Ziel-KS dagegen ist, wie schon erwähnt, weltausschnittbezogen. Bei einem Übergang zu einem anderen Weltausschnitt als dem der Abwasserüberwachung bliebe somit Zwischen-KS als oberflächenbezogenes KS erhalten, während die Bedeutungspostulate dem neuen Weltausschnitt angepaßt werden müßten.